CLAIMS

What is claimed is:

1. A method for processing a speech signal, comprising:

receiving an input speech signal;

constructing a phoneme lattice for the input speech signal;

searching the phoneme lattice to produce a likelihood score for each potential path; and

determining a processing result for the input speech signal based on the likelihood score of each potential path.

2. The method of claim 1, wherein constructing the phoneme lattice comprises:

segmenting an input speech signal into frames;

extracting acoustic features for a frame of the input speech signal;

determining K-best initial phoneme paths leading to the frame based on a first score of each potential phoneme path leading to the frame; and

calculating a second score for each of the K-best phoneme paths for the frame.

3. The method of claim 2, further comprising:

clustering together K-best initial phoneme paths for at least one consecutive frame;

selecting M-best refined phoneme paths among the clustered phoneme paths based on second scores of these paths; and

identifying vertices and arc parameters of the phoneme lattice for the input speech signal.

4. The method of claim 2, wherein the first score and the second score comprise a score based on phoneme acoustic models and language models.

5. The method of claim 1, wherein searching the phoneme lattice comprises:

receiving a phoneme lattice;

traversing the phoneme lattice via potential paths;

computing a score for a traversed path based on at least one of a phoneme confusion matrix and a plurality of language models; and

modifying the score for the traversed path.

6. The method of claim 5, wherein modifying the score comprises adjusting the score by at least one of the following: allowing repetition of phonemes and allowing flexible endpoints for phonemes in a path.

7. The method of claim 1, wherein determining the processing result comprises determining at least one of the following: at least one candidate textual representation of the input speech signal and a likelihood that the input speech signal contains targeted keywords.

8. A method for constructing a phoneme lattice for an input audio signal comprising:

segmenting the input audio signal into frames;

extracting acoustic features for a frame of the input audio signal;

determining K-best initial phoneme paths leading to the frame based on a first score of each potential phoneme path leading to the frame; and

calculating a second score for each of the K-best phoneme paths for the frame.

9. The method of claim 8, further comprising:

clustering together K-best initial phoneme paths for at least one consecutive frame;

selecting M-best refined phoneme paths among the clustered phoneme paths based on second scores of these paths; and

identifying vertices and arc parameters of the phoneme lattice for the input speech signal.

10. The method of claim 8, wherein the first score and the second score comprises a score based on phoneme acoustic models and language models.

11. A method for searching a phoneme lattice, comprises:
receiving a phoneme lattice;
traversing the phoneme lattice via potential paths; and
computing a score for a traversed path based on at least one of a phoneme confusion matrix and a plurality of language models.

12. The method of claim 11, further comprising modifying the score for the traversed path.

13. The method of claim 12, wherein modifying the score comprises adjusting the score by at least one of the following: allowing repetition of phonemes and allowing flexible endpoints for phonemes in a path.

14. The method of claim 11, further comprising determining a search result for the input audio signal based on the modified score of each searched path.

15. A method for distributing speech processing, comprising:
receiving an input speech signal by a client;
constructing a phoneme lattice for the input speech signal by the client;
transmitting the phoneme lattice from the client to a server; and
searching the phoneme lattice to produce a result for the input speech signal for the purpose of at least one of recognizing speech and spotting keywords, in the input speech signal.

16. The method of claim 15, wherein constructing the phoneme lattice comprises:

segmenting an input speech signal into frames;

extracting acoustic features for a frame of the input speech signal;

determining K-best initial phoneme paths leading to the frame based on a first score of each potential phoneme path leading to the frame; and

calculating a second score for each of the K-best phoneme paths.

17. The method of claim 16, further comprising:

clustering together K-best initial phoneme paths for at least one consecutive frame;

selecting M-best refined phoneme paths among the clustered phoneme paths based on second scores of these paths; and

identifying vertices and arc parameters of the phoneme lattice for the input speech signal.

18. The method of claim 16, wherein the first score and the second score comprise a score based on phoneme acoustic models and phoneme language models.

19. The method for claim 15, wherein searching the phoneme lattice comprises:

receiving a phoneme lattice;

traversing the phoneme lattice via potential paths;

computing a likelihood score for a traversed path based on at least a phoneme confusion matrix and a plurality of language models;

modifying the score for the traversed path; and

determining a search result for the input audio signal based on the modified score of each searched path.

20. The method of claim 19, wherein modifying the score comprises adjusting the score by at least one of the following: allowing repetition of phonemes and allowing flexible endpoints for phonemes in a path.

21. A method for training a phoneme confusion matrix, comprising:

initializing the phoneme confusion matrix;

estimating confusion probabilities between phonemes based on a training database, and the initial phoneme confusion matrix; and

updating the phoneme confusion matrix based on the estimated confusion probabilities.

22. The method of claim 21, wherein the training database comprises a plurality of utterances, actual phoneme sequences corresponding to the plurality of utterances, and time alignment information between utterances and actual phoneme sequences of the utterances.

23. The method of claim 21, wherein estimating the confusion probabilities comprises:

constructing a phoneme lattice for each utterance in the training database;

searching the phoneme lattice to produce a phoneme sequence hypothesis for the corresponding utterance; and

estimating the confusion probabilities between phonemes based on statistics obtained by comparing actual phoneme sequences and corresponding phoneme sequence hypotheses.

24. A speech processing system, comprising:

a phoneme lattice constructor to construct a phoneme lattice for an input speech signal;

a phoneme lattice search mechanism to search the phoneme lattice for the purpose of at least of recognizing speech and spotting keywords, in the input speech signal;

a plurality of models for lattice construction; and

a plurality of models for lattice search.

25. The system of claim 24, wherein the phoneme lattice constructor comprises:

an acoustic feature extractor to segment the input speech signal into frames and to extract acoustic features for a frame;

a phoneme path estimator to determine K-best initial phoneme paths leading to the frame;

a global score evaluator to determine M-best refined phoneme paths based on a cluster of K-best paths of at least one consecutive frame; and

a lattice parameter identifier to identify lattice vertices and arc parameters based on M-best refined phoneme paths of each frame.

26. The system of claim 24, wherein the plurality of models for lattice construction comprise a plurality of phoneme acoustic models and a plurality of language models.

27. The system of claim 24, wherein the plurality of models for lattice search comprise a phoneme confusion matrix and a plurality of language models.

28. A system for constructing a phoneme lattice, comprising:

an acoustic feature extractor to segment an input speech signal into frames and to extract acoustic features for a frame;

a phoneme path estimator to determine K-best initial phoneme paths leading to the frame;

a global score evaluator to determine M-best refined phoneme paths based on a cluster of K-best paths of at least one consecutive frame; and

a lattice parameter identifier to identify lattice vertices and arc parameters based on M-best refined phoneme paths of each frame.

29. The system of claim 28, wherein the phoneme path estimator comprises a likelihood score evaluator to calculate a first score for a potential phoneme path leading to each frame.

30. The system of claim 28, wherein the global score evaluator comprises a score computation component to calculate a second score for each of K-best initial phoneme paths for each frame.

31. A distributed speech processing system, comprising:

a client to receive an input speech signal and to construct a phoneme lattice for the input speech signal; and

a server to search the phoneme lattice to produce a result for the input speech signal for the purpose of at least one of recognizing speech and spotting keywords, in the input speech signal.

32. The system of claim 31, wherein the client comprises a phoneme lattice constructor to construct a phoneme lattice and a transmitting component to transmit the phoneme lattice to the server.

33. The system of claim 31, wherein the server comprises a receiving component to receive the phoneme lattice from the client and a phoneme lattice search mechanism to search the phoneme lattice.

34. A system for training a phoneme confusion matrix, comprising:

a confusion matrix initializer to initialize the phoneme confusion matrix;

a phoneme lattice constructor to construct a phoneme lattice for each utterance in a training database; and

a phoneme lattice search mechanism to search the phoneme lattice to produce a phoneme sequence hypothesis for the corresponding utterance, based on the initial phoneme confusion matrix and a plurality of language models.

35. The system of claim 34, further comprising a confusion matrix updater to update the initial phoneme confusion matrix using confusion probabilities between phonemes estimated from statistics obtained by comparing actual phoneme sequences and corresponding phoneme sequence hypotheses.

36. The system of claim 35, wherein the phoneme confusion matrix updater comprises a confusion probability estimator to estimate confusion probabilities between phonemes based on the training database.

37. An article comprising: a machine accessible medium having content stored thereon, wherein when the content is accessed by a processor, the content provides for processing a speech signal by:

receiving an input speech signal;

constructing a phoneme lattice for the input speech signal;

searching the phoneme lattice to produce a likelihood score for each potential path; and

determining a processing result for the input speech signal based on the likelihood score of each potential path.

38. The article of claim 37, wherein content for constructing the phoneme lattice comprises content for:

segmenting an input speech signal into frames;

extracting acoustic features for a frame of the input speech signal;

determining K-best initial phoneme paths leading to the frame based on a first score of each potential phoneme path leading to the frame; and

calculating a second score for each of the K-best phoneme paths for the frame.

39. The article of claim 38, further comprising content for:

clustering together K-best initial phoneme paths for at least one consecutive frame;

selecting M-best refined phoneme paths among the clustered phoneme paths based on second scores of these paths; and

identifying vertices and arc parameters of the phoneme lattice for the input speech signal.

40. The article of claim 38, wherein the first score and the second score comprise a score based on phoneme acoustic models and language models.

41. The article of claim 37, wherein content for searching the phoneme lattice *comprises content for*:

receiving a phoneme lattice;

traversing the phoneme lattice via potential paths;

computing a score for a traversed path based on at least one of a phoneme confusion matrix and a plurality of language models; and

modifying the score for the traversed path.

42. The article of claim 41, wherein content for modifying the score comprises content for adjusting the score by at least one of the following: allowing repetition of phonemes and allowing flexible endpoints for phonemes in a path.

43. The article of claim 37, wherein content for determining the processing result comprises content for determining at least one of the following:

at least one candidate textual representation of the input speech signal and a likelihood that the input speech signal contains targeted keywords.

44. An article comprising: a machine accessible medium having content stored thereon, wherein when the content is accessed by a processor, the content provides for constructing a phoneme lattice for an input audio signal by:

segmenting the input audio signal into frames;

extracting acoustic features for a frame of the input audio signal;

determining K-best initial phoneme paths leading to the frame based on a first score of each potential phoneme path leading to the frame; and

calculating a second score for each of the K-best phoneme paths for the frame.

45. The article of claim 44, further comprising content for:

clustering together K-best initial phoneme paths for at least one consecutive frame;

selecting M-best refined phoneme paths among the clustered phoneme paths based on second scores of these paths; and

identifying vertices and arc parameters of the phoneme lattice for the input speech signal.

46. The article of claim 44, wherein the first score and the second score comprises a score based on phoneme acoustic models and language models.

47. An article comprising: a machine accessible medium having content stored thereon, wherein when the content is accessed by a processor, the content provides for searching a phoneme lattice by:

receiving a phoneme lattice;

traversing the phoneme lattice via potential paths; and

computing a score for a traversed path based on at least one of a phoneme confusion matrix and a plurality of language models.

48. The article of claim 47, further comprising content for modifying the score for the traversed path.

49. The article of claim 48, wherein content for modifying the score comprises content for adjusting the score by at least one of the following: allowing repetition of phonemes and allowing flexible endpoints for phonemes in a path.

50. The article of claim 47, further comprising content for determining a search result for the input audio signal based on the modified score of each searched path.

51. An article comprising: a machine accessible medium having content stored thereon, wherein when the content is accessed by a processor, the content provides for distributing speech processing by:

receiving an input speech signal by a client;

constructing a phoneme lattice for the input speech signal by the client;

transmitting the phoneme lattice from the client to a server; and

searching the phoneme lattice to produce a result for the input speech signal for the purpose of at least one of recognizing speech and spotting keywords, in the input speech signal.

52. The article of claim 51, wherein content for constructing the phoneme lattice comprises content for:

segmenting an input speech signal into frames;

extracting acoustic features for a frame of the input speech signal;

determining K-best initial phoneme paths leading to the frame based on a first score of each potential phoneme path leading to the frame; and

calculating a second score for each of the K-best phoneme paths.

53. The article of claim 52, further comprising content for:

clustering together K-best initial phoneme paths for at least one consecutive frame;

selecting M-best refined phoneme paths among the clustered phoneme paths based on second scores of these paths; and

identifying vertices and arc parameters of the phoneme lattice for the input speech signal.

54. The article of claim 52, wherein the first score and the second score comprise a score based on phoneme acoustic models and phoneme language models.

55. The article for claim 51, wherein content for searching the phoneme lattice comprises content for:

receiving a phoneme lattice;

traversing the phoneme lattice via potential paths;

computing a likelihood score for a traversed path based on at least a phoneme confusion matrix and a plurality of language models;

modifying the score for the traversed path; and

determining a search result for the input audio signal based on the modified score of each searched path.

56. The article of claim 55, wherein content for modifying the score comprises content for adjusting the score by at least one of the following: allowing repetition of phonemes and allowing flexible endpoints for phonemes in a path.

57. An article comprising: a machine accessible medium having content stored thereon, wherein when the content is accessed by a processor, the content provides for training a phoneme confusion matrix by:

initializing the phoneme confusion matrix;

estimating confusion probabilities between phonemes based on a training database, and the initial phoneme confusion matrix; and

updating the phoneme confusion matrix based on the estimated confusion probabilities.

58. The article of claim 57, wherein the training database comprises a plurality of utterances, actual phoneme sequences corresponding to the plurality of utterances, and time alignment information between utterances and actual phoneme sequences of the utterances.

59. The article of claim 57, wherein content for estimating the confusion probabilities comprises content for:

constructing a phoneme lattice for each utterance in the training database;

searching the phoneme lattice to produce a phoneme sequence hypothesis for the corresponding utterance; and

estimating the confusion probabilities between phonemes based on statistics obtained by comparing actual phoneme sequences and corresponding phoneme sequence hypotheses.